

# Human Motion Segmentation using Active Shape Models

**Maria João M. Vasconcelos and João Manuel R. S. Tavares**

Instituto de Engenharia Mecânica e Gestão Industrial,  
Faculdade de Engenharia, Universidade do Porto  
Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal  
[maria.vasconcelos, tavares}@fe.up.pt](mailto:{maria.vasconcelos, tavares}@fe.up.pt)

**Abstract** Human motion analysis from images is meticulously related to the development of computational techniques capable of automatically identifying, tracking and analyzing relevant structures of the body. This work explores the identification of such structures in images, which is the first step of any computational system designed to analyze human motion. A widely used database (CASIA Gait Database) was used to build a Point Distribution Model (PDM) of the structure of the human body. The training dataset was composed of 14 subjects walking in four directions, and each shape was represented by a set of 113 labelled landmark points. These points were composed of 100 contour points automatically extracted from the silhouette combined with an additional 13 anatomical points from elbows, knees and feet manually annotated. The PDM was later used in the construction of an Active Shape Model, which combines the shape model with grey level profiles, in order to segment the modelled human body in new images. The experiments with this segmentation technique revealed very encouraging results as it was able to gather the necessary data of subjects walking in different directions using just one segmentation model.

## Introduction

Human motion analysis from images is strictly related to the development of computational techniques capable of automatically identifying, tracking and analyzing relevant structures of the human body. Many systems designed for this purpose begin with the feature extraction task, which is to identify the structures to be analyzed in the image sequences, and then followed by the establishment of feature correspondences, which is to address the problem of matching features from two consecutive frames, and, finalizing with high level analysis of human movements and poses.

Most of the methods proposed for human motion analysis until now have used models to fit body parts to the input images. The human body structure can therefore, be represented by means of stick figures, 2D contours or volumetric models.

While stick figures are mainly built to learn the 3D variability of the human posture and perform gait detection, as in [1,2]; 2D contours, instead, are often extracted to perform human detection and tracking [3,4]; on the other hand, volumetric models have many potential applications, including advanced Human Computer Interaction, 3D animation and robot control [5]. Therefore, in this work, we intended to build a model combining information from the stick figures and from the 2D silhouette contours in order to obtain the advantages of both modelling methodologies.

Active Shape Models (ASMs) [6] have commonly been used in different domains of Computational Vision for image segmentation purposes, in areas such as medicine, industry and security. The aim here was to build a Point Distribution Model (PDM) [7] that combines the human silhouette contour and the specific anatomical joints from the stick figure and then to construct an ASM in order to segment the modelled body further into new images.

In [8] the authors proposed a model-based on a pose of the human body using ASMs with stick figures for frontal poses and they obtained good results. However, their research did not take into account the influence of illumination and was restricted to one view of the human pose. Recently, [9] introduced Active Shape Feedback Segmentation, which is based on the Grab-cut segmentation framework and ASMs, but the authors claim that there were still some problems to be solved to be able to use the system in real situations. Also, a small dataset was used, and the time required was a drawback.

A rich collection of gait datasets have been collected, which the research community has been using for comparative performance evaluations. Examples of those databases include CMU Mobo, HID-UMD datasets, SOTON databases, HumanID Gait Challenge, CASIA databases and TUM GAID [10]. A description of the existing gait databases through to 2010 can be found in [11].

There are various applications for human motion analysis. Clinical studies, for instance, require accurate motion knowledge for the diagnosis of locomotion difficulties or abnormalities in patients. Another application is in sports, where it is commonly used to find potential improvements in an athlete's performance.

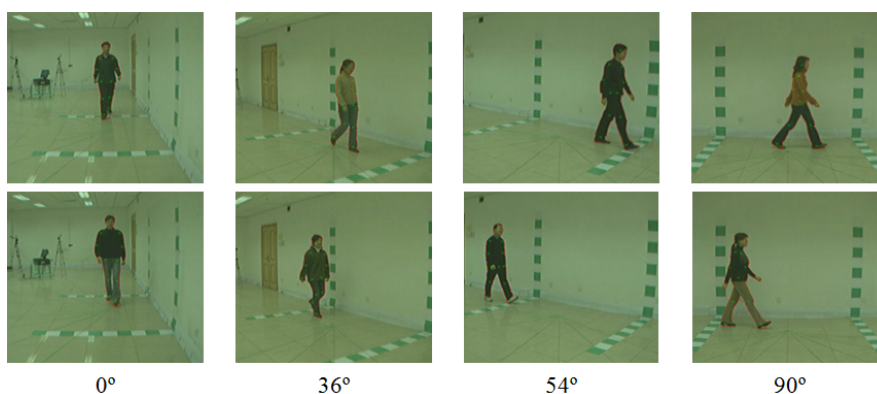
This chapter is organized as follows: in the next section, the methodology adopted is presented, starting by describing the image database used and also explaining the construction of Point Distribution Models and Active Shape Models. Afterwards, the model built and their application in the segmentation of the human body in new images are presented, and the results are shown and discussed. Finally, the conclusions are made and some considerations for future work are pointed out.

## Materials & Methods

### *Database*

The CASIA Gait Database [12], which is a large multiview gait database created in January 2005 and widely adopted by the Computational Vision community, was used here in order to test the efficiency of the models built. In addition to the original video files, the referred database also contains the human silhouettes extracted from them. All video sequences were stored as video files encoded with MJPEG, a frame rate of 25 fps and a frame size of 320x240 pixels. The database also has information about the 124 subjects involved, which were mostly young Asians aged between 20 and 30.

In this work information from 14 subjects walking in four different directions ( $0^\circ$ ,  $36^\circ$ ,  $54^\circ$  and  $90^\circ$ ) in relation to the camera was used (Figure 1). Table 1 depicts the number of images used in the construction of the models built, as well as the number of images used to test their segmentation accuracy.



**Fig. 1** Examples of subjects walking in different directions (images used for training in the first row, and testing in the second row).

**Table 1** Summary of the data used to construct (training) and test the segmentation models built.

Direction	# training images	# testing images
00	746	73
36	696	52
54	695	65
90	597	25
Total	2734	215

### ***Point Distribution Models and Active Shape Models***

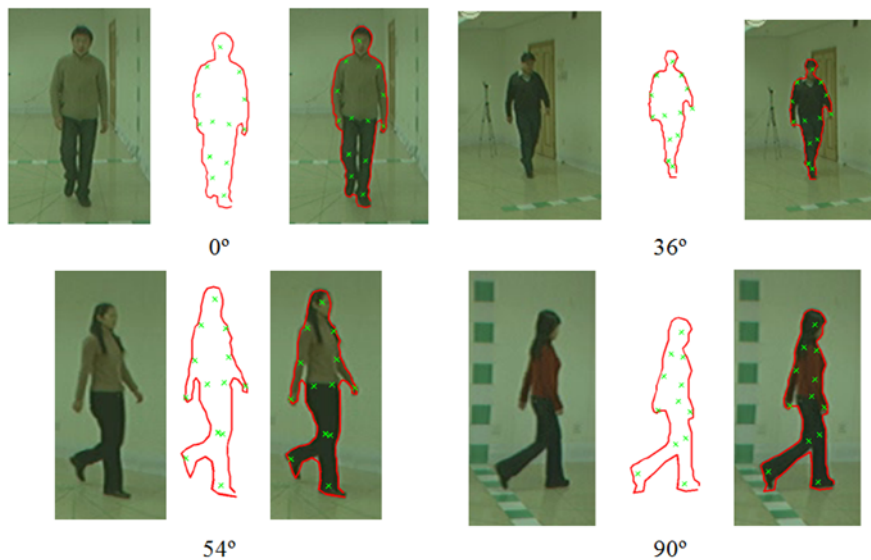
A Point Distribution Model expresses the mean shape of the modelled object in addition to the admissible variations in relation to the same mean shape [7]. In this work, the silhouette contour and the anatomical landmark points of the human body structure were modelled by a PDM from a set of 2734 images. Those images include various configurations of the human body acquired while different persons were walking (Figure 1).

In order to obtain a robust PDM, the images used in the training process needed to adequately represent the variability of the human shape during walking. Moreover, each human silhouette presented in the training set had to be described by a group of labelled landmark points conveying important aspects of the body contour. Hence, 100 contour points were chosen to be extracted from the silhouettes available on the dataset and a further 13 extra points indicating the anatomical points from the human stick figure were manually selected, leading to a total of 113 landmark points to represent the structure of the human body. In more detail, the landmark points used were the following:

- From the silhouette contour:
  - o 45 points from the left side (equally spaced);
  - o 45 points from the right side (equally spaced);
  - o 10 points between the feet of the subject (equally spaced);
- From the stick figure:
  - o 1 point in the centre of the head (1);
  - o 1 point on each shoulder (2);
  - o 1 point on each elbow (2);
  - o 1 point on each hand (2);
  - o 1 point on left and right of the hip (2);
  - o 1 point on each knee (2);
  - o 1 point on the backside of each foot (2).

In all the images here, the landmark points corresponding to the silhouette contour appear connected by fictitious line segments to enhance their visualization, while the anatomical landmark points are represented with an “x”. Figure 2 shows examples of some images used with the corresponding extracted landmark points.

In order to learn the variability of the coordinates of the landmark points in the training shapes, the points need to be aligned using, for example, dynamic programming. Only afterwards, it is possible to calculate the mean shape and apply the Principal Component Analysis (PCA) to the deviations from the mean shape and obtain the modes of variation, which characterize the manner in which the landmarks of the modelled object tend to move together. Once the PDM of the object under study is built, it can be further used to generate new shapes of the object as well as to segment it in new images through an Active Shape Model developed for the same object as will be explained next.



**Fig. 2** The landmark points used to build the models in an example image of each direction studied.

Similarly to the coordinates of the landmark points, the local gray level surroundings of each landmark can be considered in the statistical modelling of objects from images. Thus, statistical information about the mean and covariance of the gray level values of the pixels around each landmark can be used in the object modelling.

The PDM and the gray level profiles of each landmark point used in the construction of the PDM may be used to segment the object modelled into new images through an Active Shape Model [6]. ASMs use an interactive technique that fits flexible models to objects represented in images based on an optimization scheme, which is combined with PDMs to refine the initial estimated shapes of the objects to be segmented. The refining process adopted can be summarized into following steps: 1) The displacement required to move the model to a more suitable position, that is, closer to the final shape, is calculated at each landmark point; 2) The calculation of the changes in the overall shape position, orientation and scale that most effectively satisfy the local displacements found in 1); 3) Make the required adjustments in the parameters of the model, by analyzing the residual differences between the current shape of the model and the final shape sought.

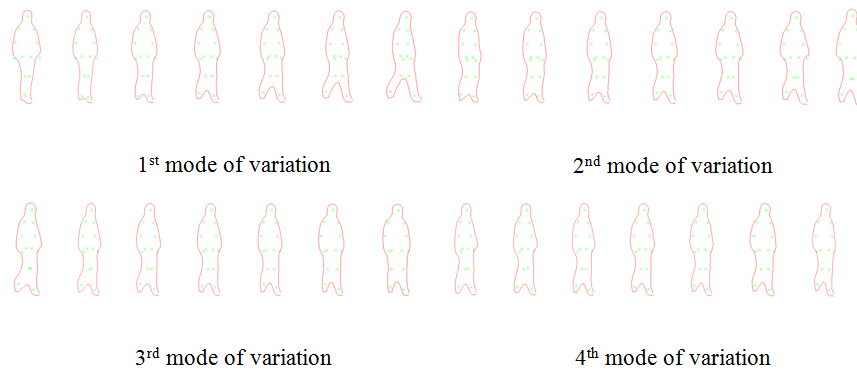
The image segmentation process with the support of ASMs was improved in [13] by adopting a multiresolution approach. In this approach, a multiresolution pyramid of the input images is built by applying a Gaussian mask; then, the grey level profiles at the various levels of the pyramid built are studied. As such, the former process enables the ASMs to segment the input images more quickly and in a reliable manner.

## Results

To extract the landmark points from the silhouettes on CASIA Gait Database, an algorithm in MATLAB was developed to automatically identify and extract the coordinates of the 100 contour points according to the rule described in the previous section. Afterwards, the 13 anatomical landmark points corresponding to the stick figure of the human body were manually extracted from each frame, and the associated coordinates were concatenated with the contour related landmarks.

The Active Shape Model software [14] was used for the ASM building process. In this process 95% of all object shape variance in the geometrical modelling (i.e. in PDM) was adopted and a 7 pixels width profile was used for the grey level modelling. As a stopping criterion for the segmentation process, a maximum of 6 iterations for each resolution level was applied. As 3 resolution levels were defined based on the dimensions of the images under study, only 18 iterations were performed. Finally, for the segmentation quality assessment, the mean and standard deviation values of the Euclidean distances between the landmark points of the final shape and the ground-truth reference points were calculated.

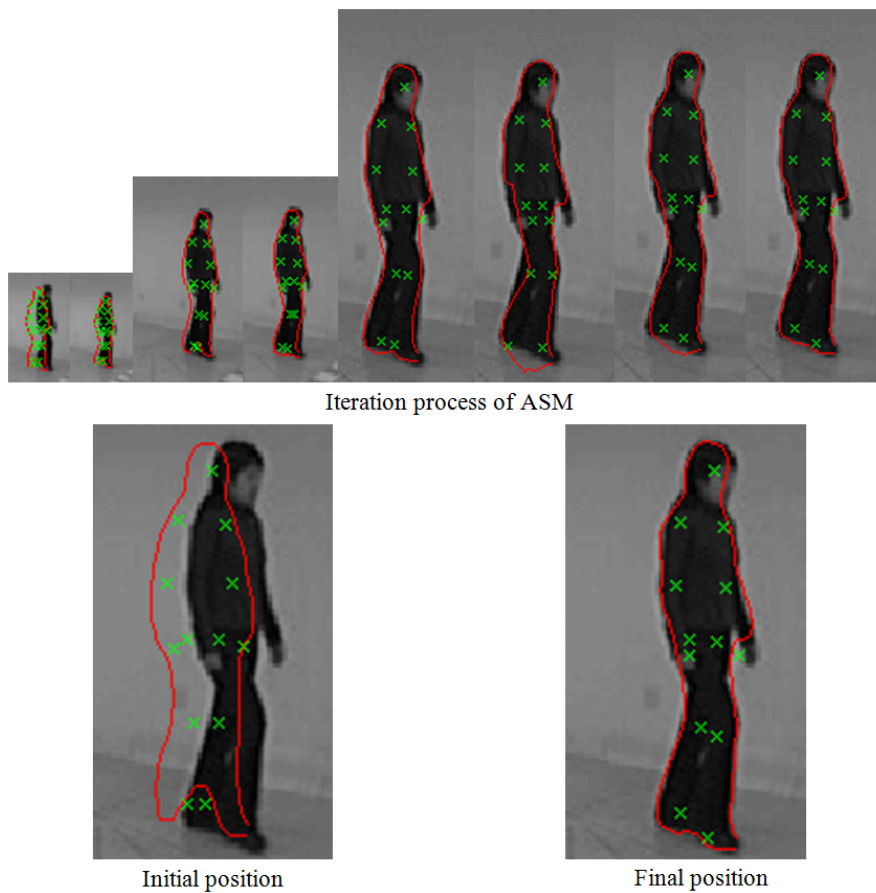
The adequacy of the modelling process to characterizing the human shape is shown by the more significant modes of variation of the PDM built in Fig 3. For instance, it is clear that the first mode of variation gathers the information on the walking stance of the subjects, whilst the second and third modes of variation gather the direction in which the subject is walking. In contrast, smaller and more specific variations can be seen in the fourth mode.



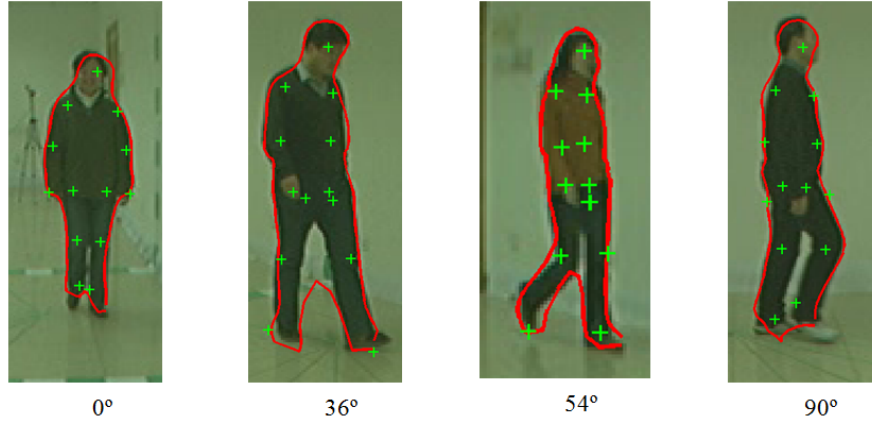
**Fig. 3** First four modes of variation of the PDM built (mean shape  $\pm$  1 standard deviation).

Figure 4 depicts several steps of the active search performed in order to segment an image from the test set; not included in the training set used, and the adaptation of the ASM built along the iteration process to reach an optimal result can be seen. Other examples, in this case, just showing the final position for the 4 directions studied are represented in Figure 5. These images show that the land-

marks corresponding to contour points have a much more reliable behaviour than those corresponding to the anatomical points. In fact, this behaviour was expected since the ASM searches for the grey level information around the point positions, and the anatomical landmark points are more likely to have similar neighbours around them, making it more difficult to choose the right position in comparison to the landmark contour points.



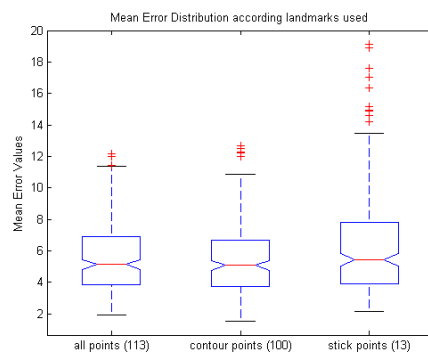
**Fig. 4** Example of the iteration process using an active shape model in a new image in the first row, and, in the second row, the initial and final (i.e. the segmentation result) positions of the model.



**Fig. 5** Examples of segmentation results obtained in images for the 4 directions studied.

In order to conclude on the variation errors of the model, besides the mean Euclidean error distribution calculation for all 113 points, one also decided to study the mean Euclidean error distribution for each subgroup of points: the contour landmarks and the anatomical landmarks (i.e. from the stick figure), separately. In other words, the error distribution was calculated using 113, 100 and 13 points, corresponding to the all the points, the contour points and the stick model.

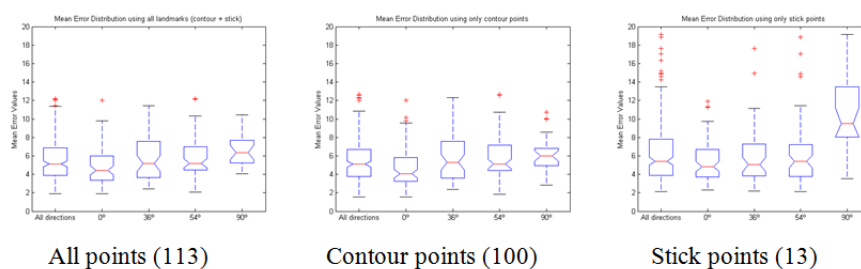
As expected, from the observation of data distribution obtained, see Figure 6, one can confirm that the mean error distribution is slightly worse for the subgroup of the stick points. If we take into account that the images under study have 320x240 pixels in size, it is worth noting that the results achieved with the suggested segmentation model are extremely satisfactory, within the 25<sup>th</sup> to 75<sup>th</sup> percentile interval ranging from 4 to 7 pixels, which translates into very accurate segmentation results. Even the mean error distribution considering only the stick points achieves good results, with the 25<sup>th</sup> to 75<sup>th</sup> percentile ranging from 4 to 8 pixels.



**Fig. 6** Mean error distribution according to the landmark point set used.



We also studied the segmentation quality in terms of the direction in which the subjects were walking, Figure 7. Analyzing this figure, an interesting conclusion can be reached: considering all directions against each direction separately, equivalent error ranges were achieved. This behaviour is repeated independently of the landmark points. In summary, the ASM built could successfully segment the human body structure in images independently of the direction the subjects were walking, which is an important goal in studies concerning human motion analysis from images.



**Fig. 7** Mean error distributions according to the direction of the subjects and using all the 113 landmark points; only the landmark points from the contour; or just the anatomical landmark points.

## Conclusions

In this work, a widely used database, CASIA Gait Database, was used to build the Point Distribution Model of the structure of the human body. The training image dataset used was composed of 14 subjects walking in four directions, and each shape of the training set was represented by a set of 113 labelled landmark points, which combines information of the contour of the silhouette and the anatomical stick points.

In order to obtain the mean shape of the human silhouette as well as its admissible shape variations a PDM was built. This model was then used to construct an Active Shape Model, which combines the shape model (i.e. PDM) with grey level profiles, with the purpose of segmenting the modelled silhouettes into new images.

The good results obtained through the use of the ASM built to perform human body segmentation in new images strongly suggest that this type of deformable model can be used in this task. In addition, just one segmentation model gathers the necessary information to segment the structure of the human body independently of the direction the subjects were walking.

A future work would be to develop methodologies that can combine, more accurately, the human silhouette with important anatomical joint positions, mainly

for use in biomechanical studies related to human motion. As soon as the use of simple image cameras allows the robust and detailed analysis of real movements performed by subjects in their daily life, it will be possible to obtain new levels of information of the subjects from the input images. Up until now, this has been only achieved from images acquired under well controlled conditions and in significantly restricted environments, which consequently demands more robust techniques of image segmentation, motion tracking and analysis.

**Acknowledgments** The first author would like to thank the support of the PhD grant with references SFRH/BD/28817/2006 from *Fundação para a Ciência e Tecnologia* (FCT), in Portugal. This work was partially developed under the scope of the project with reference PTDC/BBB-BMD/3088/2012 financially supported by FCT. The research described in this chapter uses the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences.

## References

1. Ogawara K, Li X, Ikeuchi K. Marker-less Human Motion Estimation using Articulated Deformable Model. IEEE International Conference on Robotics and Automation, 2007. p. 46-51.
2. Wei XK, Chai J. Intuitive Interactive Human-Character Posing with Millions of Example Poses. IEEE Computer Graphics Applications, 2011. 31(4): p. 78-88.
3. Al-Huseiny M, Mahmoodi S, Nixon M. Gait Sequence Synthesis and Reconstruction. IEEE Transactions on PAMI 2008. 30.8: p. 1385-1399.
4. Das Choudhury S, Tjahjadi T. Gait recognition based on shape and motion analysis of silhouette contours. Computer Vision and Image Understanding, 2013. 117(12): 1770-85.
5. Tran C, Trivedi MM. Human body modelling and tracking using volumetric representation: Selected recent studies and possibilities for extensions. Second ACM/IEEE International Conference on Distributed Smart Cameras, 2008. p. 1-9.
6. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models-Their Training and Application. Computer Vision and Image Understanding, 1995. 61(1): p. 38-59.
7. Cootes TF, Taylor CJ, Cooper DH, Graham J. Training Models of Shape from Sets of Examples. BMVC92. Springer London, 1992. p. 9-18.
8. Jang C, Jung K. Human pose estimation using Active Shape Models. Proceedings of World Academy of Science: Engineering & Technology 46 (2008).
9. Pourjam E, Ide I, Deguchi D, Murase H. Segmentation of Human Instances Using Grab-cut and Active Shape Model Feedback. Proceedings of IAPR Conference on Machine Vision Applications (MVA) 2013. p. 77-80.
10. Hofmann M, Geiger J, Bachmann S, Schuller B, Rigoll G. The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits. Journal of Visual Communication and Image Representation 2014. 25(1): p. 195-206.
11. Nixon MS, Tan T, Chellappa R. Human Identification Based on Gait. Vol 4, Springer; 2010.
12. Yu S, Tan D, Tan T. A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. 18th International Conference on Pattern Recognition, 2006. p. 441-444.
13. Cootes TF, Taylor CJ, Lanitis A. Multi-resolution search with active shape models. Proceedings of the 12th IAPR International Conference on Pattern Recognition, 1994. Vol.1 p. 610-612.

14. Hamarneh G. Active Shape Model Software [Internet]. 1999. Available from: <http://www.cs.sfu.ca/~hamarned/software/code/asm.zip>. Last accessed June 2012.